

dbt 101

A tutorial training

Agenda

- 1 Why and what is dbt
- 2 Setup review
- 3 Data warehouse with dbt
- 4 Feedback, Links, etc.

What is dbt

- data build tool - software and data warehouse development framework by dbt Labs
- It is the T in ETL - a tool to do the transformation of data in a data warehouse
- A way of declarative data model development, solely based on SQL SELECT Statement and a bunch of configuration files
- A way to standardize the definition of KPIs and metrics throughout a scope (company?)
- Written in python, hence it is available as a python package.
- Not a data loading tool or an orchestrator
- Not a database, but available for many databases - even for duckDB
- Not a tool to develop non analytical databases for application backends

Why use dbt?

- **Build and model.** You can use SQL to define transformations to apply to your source data and run these transformations in a correct order.
- **Ensure quality.** Data warehouses create consistency among different data types from disparate sources.
- **Standardize transformations.** Macros are a great way to build the same transformations every in your code. Metrics allow you to standardize business logic
- **Collaborative Data Teams.** Data warehouse tend to be one man shows and DBA are sometimes the lonely wolfs off software development. With dbt you get an environment to build data warehouses together as a team.

How to actually dbt?

Useful links

Companion script: <https://github.com/fengineering/dbt-101>

Official dbt documentation: <https://docs.getdbt.com/docs/collaborate/documentation>

dbt Synapse documentation: <https://docs.getdbt.com/reference/warehouse-setups/azuresynapse-setup>

dbt Sql Server: <https://github.com/dbt-msft/dbt-sqlserver>
